

The University of Chicago

A TEST OF
STUDENTS' COMPETENCE TO USE
THE LIBRARY

A PART OF A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE LIBRARY
SCHOOL IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

1937

By
LULU RUTH REED

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

Reprinted from
THE LIBRARY QUARTERLY
Vol. VIII, No. 2, April 1938

The University of Chicago

A TEST OF
STUDENTS' COMPETENCE TO USE
THE LIBRARY

A PART OF A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE LIBRARY
SCHOOL IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

1937

By
LULU RUTH REED

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

Reprinted from
THE LIBRARY QUARTERLY
Vol. VIII, No. 2, April 1938

A TEST OF STUDENTS' COMPETENCE TO USE THE LIBRARY¹

LULU RUTH REED

THE attitude of both librarians and educators toward the problem of student competence in the use of library facilities is paradoxical. Professional literature abounds with suggestions about instruction in the use of the library, and yet comparatively little has been done through definitely planned courses. The situation was set forth convincingly in a report made by the Office of Education in 1932. Here it was pointed out that librarians and educators alike agreed that high school students should be given instruction in the use of books and the library, but that not more than 31 per cent of the 766 schools studied gave any instruction and that the percentage for the country as a whole was probably much lower because of the failure of many of the schools to reply to the questionnaires.² That educators are becoming increasingly interested in a better understanding of study habits and skills is evident from the procedures now being developed by the Eight Year Study under the auspices of the Progressive Education Association.³ This situation, and particularly the attitude of educators, may be accounted for, in part, by the lack of adequate measuring instruments which would make it possible to relate student competence in the use of the library to skill in other fields.

The comparatively recent origin of scientific measurement in the history of examinations is rather surprising in view of the simplicity of the idea basic to standardized, objective tests. In

¹ Unpublished Doctor's thesis, Graduate Library School, University of Chicago, 1937.

² B. L. Johnson, *The secondary-school library* (U.S. Office of Education *Bulletin* No. 17; "Monograph" No. 17; Washington: Government Printing Office, 1932), p. 43.

³ R. W. Tyler, "The study of adolescent reading by the Progressive Education Association," *Library trends*, ed. by Louis R. Wilson (Chicago: University of Chicago Press, 1937), p. 272.



a recent publication on the use of tests, Greene and Jorgensen point out very clearly the principle involved. The purpose is stated as the selection and administration of typical exercises to groups of individuals until there is a norm or standard set up, showing the typical reactions of individuals to particular subjects.⁴

Since there were no adequately standardized tests available for the investigation of differences in student competence in the use of the library, it was necessary to devise such tests before the relationship between library skills and other study habits and skills could be investigated. Tests were constructed, utilizing the methods and procedures provided by experts in the field of educational measurement. Although all sciences are alike in the fundamental principles of measurement—the purpose being to measure objectively qualities, quantities, and relationships—the techniques for the determination of school achievement and for the diagnosing of educational difficulties are most nearly akin to the problem of ascertaining student ability to use the library. Just as the teacher wishes to know the types of difficulties which the student of an academic subject is encountering and the extent to which he has mastered the content of the course, so the librarian wishes to understand just what it is in the utilization of library tools and books that is causing trouble and to what extent the student has achieved independent skill in their use.

THE PURPOSE OF THE STUDY

After the tests were developed it was necessary to determine whether they were reliable and valid measures for the investigation of differences in ability to utilize library resources. The primary object of this study is to show the reliability and validity of the tests; the secondary object is to set forth such differences in student competence as are shown by the scores obtained during the process of determining the reliability of the tests. The differences discussed are those that could be deduced

⁴ H. A. Greene and A. N. Jorgensen, *The use and interpretation of high school tests* (New York: Longmans, Green, 1936), p. 2.

from the scores on the tests as a whole, and on the major divisions of which there are seven in the college forms, five in the high school, and three in the elementary. No attempt is made to take up the questions that can be answered only by an item tabulation of test results, which would give the scores for each question; but the possibilities for additional studies on the data collected are indicated and the problems outlined. Nor is any attempt made to include a consideration of the effect that present library organization may have on the successes or failures of students in using the library. This is left for later study. The examination of student differences may contribute both to the study of library organization and to the consideration of remedial measures for increasing student competence in the use of the library.

Although the purpose of the study might be extended to include the specific measures to be used after difficulties have been determined, their exclusion is based on the prevalent idea in educational literature that tests should be regarded as educational barometers and that they do not provide in themselves remedial measures but must be followed by constructive supervisory programs. Another limitation to the scope of the study is the exclusion of any attempt to demonstrate that purely diagnostic tests, rather than achievement tests with diagnostic value, have been constructed.

THE DEVELOPMENT OF THE LIBRARY TESTS

Preliminary to the construction of the library tests, interviews were held with the reference librarians of the University of Chicago Library, with the faculty of the Graduate Library School, and with selected students in the school. They were asked to indicate the traits and skills necessary for successful utilization of library resources and to evaluate in terms of personal experience the usefulness of the instruction they had received in the use of the library.

Abilities involved in the maximum utilization of library resources.—The points outlined on the basis of interviews and consideration of library literature were:

1. Ability to find departments of a library
2. Ability to recognize differences in purpose of the various departments of a library and the type of service rendered by each to the library patron
3. Ability to use library material without infringing on the rights of others
4. Ability to ascertain whether a library contains a specific book
5. Ability to locate books illustrated, edited, or translated by individuals or organizations other than authors
6. Ability to ascertain whether a library contains books on specific subjects
7. Ability to find parts of books brought out by the dictionary card catalog
8. Ability to find material in books, as distinct from general subjects brought out by the dictionary card catalog. This includes a number of subdivisions:
 - a) Use of indexes and other parts of books to locate specific information
 - b) Use of special indexes, such as Firkins' *Index to short stories*
 - c) Knowledge of the arrangement and special features of comprehensive reference books such as encyclopedias and dictionaries
 - d) The use of the most specific book on a topic
 - e) Understanding of subject matter read
9. Ability to find material in magazines. This includes:
 - a) Use of contents and volume indexes
 - b) Use of general magazine indexes such as *Readers' guide*
10. Ability to find material in newspapers. This includes:
 - a) Use of current, nonindexed papers
 - b) Use of newspaper indexes, such as the *New York times index*
11. Ability to find material in government documents. This includes:
 - a) Understanding of corporate entries in the card catalog
 - b) Use of general magazine indexes to find document material
 - c) Use of special indexes to government documents
12. Ability to distinguish degrees of authenticity in material.

In the final forms of the tests there are no sections covering items 1, 3, 8e, 9a, and 11a. The omissions are due, in part, to the length of the tests, and, in part, to the fact that the same type of information could be secured in other ways, such as the use of a reading comprehension test to cover 8e.

Differences between preliminary and revised tests.—Based on these principles, a preliminary test was constructed, and administered to small groups of high school Freshmen and Seniors at the University of Chicago; to practically all of the summer students at the State Teachers College, Valley City, North Dakota; and to a small group at Central State Teachers College at Mt. Pleasant, Michigan. Returns were studied and tests

were revised separately for college, high school, and elementary school, with two forms for each level.

The general differences between the preliminary and the final forms are summarized below.

1. *Expansion of the number of items.*—The section on the use of the catalog was expanded most. The changes prevent chance successes from affecting the scores seriously. Although the titles should not affect the responses, since all information necessary for the choice is given, the additional precaution was taken of selecting the material from the *Children's catalog* for the elementary level, from the *Standard catalog for high school libraries* for high school, and from the *A.L.A. catalog* for college.

2. *Elimination of sections.*—Two types of eliminations were made. One was in the total omission of sections, such as that on the finding of government material through general magazine indexes. The other type of elimination was the omission of some sections from tests for one level while including them in others. This was done partly on logical grounds, partly on the data from courses of study examined, and partly on the basis of an item tabulation for small groups of Freshman and Senior high school students and Seniors in college.

That the revision was successful in bringing the tests nearer into line with the ability of students, although sufficiently difficult to prevent any student's going off the scale by scoring perfectly, is indicated by the comparison of mean scores on the preliminary and revised tests. Since the number of questions is not the same in the two tests, this is best stated in terms of the percentage of the total possible score represented by the mean scores. This appears in Table 1. Although it was not possible to give the preliminary and revised tests to the same students with the exception of a very small group of Freshmen at the University of Chicago High School, comparable groups were used.

3. *Revision of wording and of types of exercises.*—Some further changes were made in the interest of clarity. The exercises most changed were those of the section on the relation of reference questions to classification numbers. Since the primary purpose of this section was to determine how able students were to

utilize classification numbers to find topics, the decision was made to change the section in such a way as to eliminate the importance of memorizing numbers. Consequently, in the revised test the section was made into an association exercise, where the classification numbers are given and the topics listed below, the problem being to indicate in which division of the classification the student thought that material might be located. The increase in the score from almost nothing to about 50 per cent of the possible score indicates that the change was effective.

TABLE 1
COMPARISON OF MEAN SCORES ON PRELIMINARY AND
REVISED LIBRARY TESTS*

CLASS	PRELIMINARY TEST		REVISED COLLEGE TEST		REVISED HIGH SCHOOL TEST	
	College	High School	Form A	Form B	Form A†	Form B
Freshmen.....	35.32	24.31	42.03	41.33	32.51	32.17
Sophomores.....	36.17		44.36	43.20	40.60	39.21
Juniors.....	39.08		45.07	42.25	41.42	40.96
Seniors.....	39.49	38.93	46.68	43.86	41.29	45.09

* Mean scores in terms of percentages of total possible score.

† Figured without the University of Chicago High School in order to make forms comparable. Only Form A was administered at the University school. With returns from this school, means are about 2 points higher for Freshmen and Sophomores and 4 to 5 points higher for Juniors and Seniors.

Other changes included a reversal of order of choice and the use of multiple-response form instead of matching exercises. An example of the latter is shown by the section on the *Readers' guide*, which was changed in format rather than in content, with the aim of shortening the time required for interpretation of items and of making the exercise clearer. That the revision was helpful is shown by the fact that the mean score for college Seniors on the preliminary test was 3.48 and on the revised test was 5.0 for Form A and 4.68 for Form B. This rise in average performance would indicate that the form of the exercise did affect the response to some extent.

4. *Rearrangement of questions.*—Some item tabulations were made to aid in rearrangement. Correlations between the results for high school students and for Seniors in college determined

what major difference, if any, would have to be considered in reorganization. The correlations were so high as to suggest that the same order of questions would be satisfactory on both levels. Between the high school Freshmen and the total the correlation was $.86 \pm .015$; between the Seniors and the total it was $.99 \pm .001$; and between Freshmen and Seniors it was $.91 \pm .01$. The correlation between college Seniors and the total high school ranking was $.932 \pm .007$, indicating that the changes in order that had been made on the basis of the high school returns would also be satisfactory for the college forms.

THE RELIABILITY OF THE LIBRARY TESTS

All definitions of reliability given in educational literature emphasize the fact that consistent functioning of a test is of prime importance. The factors that may affect consistent performance can be grouped under two main divisions: the construction of the test, and the variations in human response to the test. Since experts in the field of educational measurement have established certain elements as the criteria of good tests, the first factor can be controlled to a large extent. By obtaining as nearly as possible typical groups of students the extent of influence from differences in human reaction can be minimized. In discussing the reliability of the tests, the following points will be considered: (1) objectivity; (2) length; (3) sampling of subject matter; (4) selection of schools; (5) statistical measures.

Objectivity.—Among the points emphasized in educational literature, the degree of objectivity achieved by a measuring instrument is primary. Symonds⁵ and Ruch⁶ both emphasize the importance of a format that will eliminate human judgment in the scoring of tests. Consequently, the first step taken to insure the reliability of the library tests was the utilization of the techniques developed in the new-type examinations. The type most used is the 5-response recognition question, which Ruch

⁵ P. M. Symonds, *Measurement in secondary education* (New York: Macmillan, 1934), p. 289.

⁶ G. M. Ruch, *The objective or new-type examination; an introduction to educational measurement* (Chicago: Scott, Foresman, 1929), p. 42.

has shown to be the second most reliable of five types of examinations, standing very close to the recall type, the reliability coefficients being .90 for recall and .89 for the 5-response type.

Length.—In determining the length of the tests, there were two deciding factors. The first is the fact that the longer test is likely to be more reliable because it reduces the element of chance or luck. It helps also to eliminate the factor of special skill or weakness in any particular area, which would give a misleading picture of the student's attainments if the test were confined to one phase. The second was the idea of adjusting the tests to fit as nearly as possible a class period, so that the greatest use could be made of them by the individual instructor, who could give one test at the beginning of the year to determine the relative proficiency of the group, and any particular area of weakness; and then could retest at the close of the year, or the instructional period, by using the other form. The actual number of items included was based on the time scale given by Ruch.⁷

In order to check the extent to which faulty timing of the tests might have affected the test results, the mean scores of the concluding sections and also of sections devoted to the same subject matter were compared. From them it was evident that performance on the final sections was much more nearly uniform than on sections devoted to the same subject matter in the three tests. The greatest difference in the scores on the use of the *Readers' guide* from grade to grade appears between the seventh and eighth on the elementary forms, indicating that subject matter rather than placement in the test was the chief factor in determining the score, since this section appears about the middle of the college test, about the beginning of the last third of the high school test, and at the close of the elementary test.

Sampling of subject matter.—It has been indicated that one of the reasons a longer test is more reliable is because it can cover a wider range of subject matter. The importance of the ma-

⁷ G. M. Ruch, *The improvement of the written examination* (Chicago: Scott, Foresman, 1934), p. 97.

terial determines its validity; the extent of the material influences reliability. The attempt to reduce chance variations appears in the sections devoted to the card catalog. The space devoted to such items is sufficient to include the various kinds of entries in the catalog.

It was also necessary so to sample the subject matter that no student would fail to score or would be able to score 100 per cent on the test as a whole. Failure to score means that the test is too difficult to measure his ability, and a perfect score means that it is too easy to measure it. The mean scores presented in Table 1 show that the range is adequate.

Selection of schools.—Since local factors inevitably influence individual schools, no one can be taken as wholly typical of all schools in the country. However, by selecting schools so as to guard against any definite loading of the sample and by providing counterbalances through the inclusion of different types of schools, it is possible to present mean scores with some assurance that they are truly representative. Consequently, in securing co-operating schools, the attempt was made to secure different environmental conditions. The results, in general, justified the selection made.

On the college level two different types of schools were selected: teachers' colleges, which as a group have been more interested in the problem of library instruction than others; and colleges, some being situated in small towns and some in metropolitan areas, including Chicago and Cleveland. Both of the city colleges include groups of students definitely interested in vocational education as well as in the usual liberal arts curriculum. In four of the schools there is some formal instruction, but no extensive course is given. The 1,295 students represent a useful sampling of college students, both from environmental and instructional points of view. The co-operating schools were Colorado State Teachers College, Greeley; Fenn College, Cleveland; George Williams College, Chicago; Indiana State Teachers College, Terre Haute; and Iowa State Teachers College, Cedar Falls.

On the high school level there may be an undue amount of

metropolitan influence in the selection of the schools, for the smallest city represented is Chicago Heights, a suburb of Chicago. The schools were the University of Minnesota Training School, St. Patrick's High School, Bloom Township High School in Chicago Heights, and the University of Chicago High School.

The schools represent four different types: a training school connected with a state university, a private school with a highly selected group of students, a large public school in an environment somewhat underprivileged, and a parochial school for boys. Hence, it may be assumed that the 1,201 students are fairly representative of students in secondary schools.

The elementary schools include 849 students from five different environments: a small industrial city, a suburban community adjacent to a large city, and three city systems. The schools assisting in the study are: Seneca School, Alliance, Ohio; St. John's Lutheran and Victory Park Schools, South Euclid, Ohio (through the co-operation of the public library where the students go for instruction); Atwater and Lake Bluff Schools, Milwaukee; University of Chicago Laboratory School; and University of Minnesota Junior High School. The grades included were mainly from the junior high school, though testing as low as the fifth grade was included. Most of the systems represent school organizations where library service is given by school libraries under the direction of the schools themselves; but the Ohio schools have a definite plan of co-operation with the public library.

Statistical measures.—The measure of reliability most stressed was the comparison of alternate items. It consists in correlating the scores made on one-half of the items with those made on the other half and correcting by the Spearman-Brown formula. As an additional check, one year in each level was selected for determining intercorrelations of the section scores with the total score (minus self-correlation), on one form of the tests. To equate the forms, the method used was to administer both forms to the same grade, approximately half of the students taking each form. On the elementary level, one other check was made

by testing the sixth grade of the University of Chicago Laboratory School twice. One section was given Form A and another Form B. Two weeks later the process was reversed, with no intervening instruction in the use of the library.

Table 2 presents the reliability coefficients for the library tests, arranged by class and sex. From this it is seen that no coefficient for any class is less than .775 in the college and that the highest is .915. In high school the range is from .86 to .94; and in the elementary school from .76 to .90. The correlations for the separate colleges and high schools were also figured and range as follows: for college, from .790 to .966; for high school, from .793 to .951. The sex differences are presented in Table 2. The table also shows that the fluctuations in separate classes are seldom as wide as those cited for the total college, high school, or elementary groups. The comparatively small fluctuations for groups differing in numbers indicate the stability of the coefficients.

Another evidence of the reliability of the tests is found in the comparison of their reliability coefficients with those of some standardized tests reported by Ruch and Stoddard.⁸ The library tests fall within the upper group of tests cited for high school and entering college students. The range on the college test for Freshmen (both forms) is from .845 to .883 for the group as a whole and from .86 to .94 for high schools. Even when tests listed in Ruch and Stoddard that are below .77 are excluded, the coefficients of the library test appear well toward the top of the group. The Barr diagnostic tests in American History have a coefficient of .77; four intelligence tests range from .88 to .95; the Iowa comprehension test (Grade XII) is .88; the Iowa placement (Grades XI–XII) range from .82 to .97; the Ruch-Cossman biology test (Grade X) is .82; and the Ruch-Popenoe general science test (high school classes) is .83.

A similar comparison was made for the elementary tests. The coefficients for the library tests range from .76 to .90, while those of some standardized geography, grammar, language, and

⁸ G. M. Ruch and G. D. Stoddard, *Tests and measurements in high school instruction* (Yonkers-on-Hudson, N.Y.: World Book Co., 1927), *passim*.

TABLE 2
RELIABILITY COEFFICIENTS FOR LIBRARY TESTS

CLASS AND SEX	COLLEGE TEST				HIGH SCHOOL TEST				GRADE	ELEMENTARY TEST			
	Coefficients		Number of Students		Coefficients		Number of Students			Coefficients		Number of Students	
Form A	Form B	A	B	Form A	Form B	A	B	Form A	Form B	A	B		
<i>Freshmen:</i>									5th Grade				
Total.....	.883 ± .007	.845 ± .009	464	453	.872 ± .017	.886 ± .015	90	97	.79 ± .025	.90 ± .013	106	90	
Male.....	.897 ± .011	.897 ± .011	158	158	.842 ± .023	.908 ± .014	72	77	.85 ± .025	.82 ± .031	57	51	
Female...	.866 ± .009	.832 ± .012	306	295	.959 ± .013	.908 ± .003	18	20	.70 ± .050	.91 ± .018	49	39	
<i>Sophomores:</i>									6th Grade				
Total.....	.915 ± .013	.898 ± .017	66	56	.878 ± .016	.878 ± .016	103	91	.87 ± .011	.87 ± .014	193	131	
Male.....	.953 ± .015	.960 ± .014	21	15	.963 ± .063	.858 ± .028	55	41	.87 ± .017	.84 ± .024	89	66	
Female...	.925 ± .014	.776 ± .042	45	41	.922 ± .015	.896 ± .019	48	50	.84 ± .020	.84 ± .024	104	65	
<i>Juniors:</i>									7th Grade				
Total.....	.775 ± .032	.880 ± .018	69	73	.904 ± .013	.940 ± .008	88	80	.87 ± .018	.78 ± .037	77	50	
Male.....	.904 ± .023	.888 ± .024	28	34	.941 ± .011	.920 ± .016	40	38	.72 ± .053	.82 ± .045	38	25	
Female...	.831 ± .032	.907 ± .020	41	39	.824 ± .031	.930 ± .015	48	42	.91 ± .018	.74 ± .063	39	25	
<i>Seniors:</i>									8th Grade				
Total.....	.881 ± .021	.912 ± .016	52	53	.901 ± .011	.860 ± .013	118	128	.85 ± .026	.76 ± .032	53	78	
Male.....	.930 ± .014	.894 ± .023	32	35	.900 ± .014	.860 ± .018	81	91	.69 ± .008	.82 ± .061	20	40	
Female...	.787 ± .057	.958 ± .013	20	18	.930 ± .004	.849 ± .031	37	37	.91 ± .019	.78 ± .042	33	38	

reading tests given in Ruch and Stoddard range from .58 to .98, and the New Stanford achievement tests range from .92 to .97.⁹ With the exception of the New Stanford tests which have been very thoroughly standardized, most fall between .8 and .9. Since achievement tests on the elementary level are based on traditional subject matter, so that students have received definite and systematic instruction before testing and consequently should be more uniform in their responses, the variations on the library tests are slight in comparison with the others.

TABLE 3

COMPARISON OF MEAN SCORES AND RELIABILITY COEFFICIENTS: LIBRARY TEST, TWO TESTINGS, SIXTH GRADE: UNIVERSITY OF CHICAGO LABORATORY SCHOOL

SECTIONS	TESTINGS							
	FIRST				SECOND			
	Means	P.E.	r^*	P.E.	Means	P.E.	r	P.E.
M.	26.19†	1.08	.837	.029	29.66‡	1.17	.910	.021
W.	22.74‡	0.86	.790	.047	29.52†	1.39	.940	.015

* Pearson Product Moment Correlation Coefficient, based on a split-half count of alternate items, corrected by the Spearman-Brown formula.

† Form B, Library Test.

‡ Form A, Library Test.

As an additional check on the elementary level, it was possible to give both forms of the test to a sixth-grade class. The two sections are referred to as W and M. Table 3 compares the two testings. It shows that both sections performed better on the second trial, which may have been owing in part to a better adjustment of time, since only twenty-five minutes were available for the first trial and thirty for the second. However, the difference between sections was more marked the first day, where the means differ by 3.45. The probable error of this difference is 1.38. On the second trial they vary by only .14, with a P.E. of the difference of 1.82. Since there is more improvement in the score of Section W, which took first Form A and

⁹ *Ibid.*

then Form B, than for Section M, which took Form B first and then A, there is some indication that Form A is slightly more difficult than B. However, the greater variation in the probable errors of the mean scores for Section W than for M and the greater variation in reliability coefficients tends to indicate that the differences are probably due more to the ability of the stu-

TABLE 4
TOTAL MEAN SCORES AND PROBABLE ERRORS: COLLEGE
LIBRARY TEST

CLASS AND SEX	FORM A			FORM B		
	Means	P.E.	Number of Students	Means	P.E.	Number of Students
<i>Freshmen:</i>						
Total.....	65.15	0.5	464	64.06	0.5	453
Male.....	63.06	0.9	158	63.47	0.9	158
Female.....	66.23	0.6	306	64.37	0.6	295
<i>Sophomores:</i>						
Total.....	68.76	1.3	66	66.96	1.3	56
Male.....	69.38	2.2	21	68.93	3.2	15
Female.....	68.47	1.5	45	66.24	1.8	41
<i>Juniors:</i>						
Total.....	69.86	1.4	69	65.49	1.3	73
Male.....	69.96	2.2	28	63.88	2.0	34
Female.....	69.78	1.6	41	66.90	1.7	39
<i>Seniors:</i>						
Total.....	72.35	1.6	52	67.98	1.7	53
Male.....	71.09	2.2	32	64.46	2.1	35
Female.....	74.35	1.9	20	74.83	1.6	18

dents than to the test forms. The greater stability of the performance of Section M is also indicated by the fact that the correlation between performance on the two forms of the test is $.881 \pm .026$ for M and $.79 \pm .045$ for W.

Besides the evidence of the equivalence of the forms afforded by this check in the sixth grade, it is also possible to compare the mean scores for the two forms on each level. These appear in Tables 4, 5, and 6. Since the groups used are too small to constitute thoroughly established norms, particularly for all but

college Freshmen, any weighting of the scores for the two forms is a doubtful procedure. This is indicated by two facts. The difference between the two forms for college Freshmen is 1.09, based on returns from 917 students, and for college Sophomores is 1.8, based on only 66 returns. This suggests that the variation

TABLE 5
TOTAL MEAN SCORES AND PROBABLE ERRORS: HIGH SCHOOL
LIBRARY TEST

CLASS AND SEX	FORM A			FORM B		
	Means*	P.E.	Number of Students	Means	P.E.	Number of Students
<i>Freshmen:</i>						
Total.....	43.89	1.0	90	43.43	1.1	97
Male.....	42.65	1.1	72	42.12	1.3	77
Female.....	48.83	2.0	18	48.50	1.9	20
<i>Sophomores:</i>						
Total.....	54.81	1.0	103	52.93	1.1	91
Male.....	55.44	1.4	55	53.37	1.6	41
Female.....	54.08	1.5	48	52.58	1.5	50
<i>Juniors:</i>						
Total.....	55.92	1.1	88	55.29	1.4	80
Male.....	54.05	1.8	40	54.18	2.1	38
Female.....	57.48	1.2	48	56.29	1.9	42
<i>Seniors:</i>						
Total.....	55.74	1.0	118	60.87	1.0	128
Male.....	54.30	1.2	81	58.31	1.2	91
Female.....	58.89	1.8	37	67.16	1.7	37

* Figured without the University of Chicago in order to show comparable means for Forms A and B. Only Form A was administered at the University of Chicago High School.

in senior college of 4.37 is due to other factors than numbers. When the mean for all college groups is computed, disregarding class, the difference is 1.78, with a mean of 66.59 on Form A and a mean of 64.81 on Form B, which is a second indication that weighting should be tentative, if used at all. The total means for high school and elementary school bear out this decision, the difference for high school being .79 and for elementary school .78. The high school means are: Form A, 52.86; Form B, 53.65. Those for elementary school are: Form A, 27.8; Form B, 28.58.

Tables 4 and 5 suggest that sex differences play an important part in the variations in college and high school. Sex variations are much less marked in the elementary school. In the Senior year in college, where the difference on total scores between the two forms is 4.37, the divergence of women on the two is only

TABLE 6
TOTAL MEAN SCORES AND PROBABLE ERRORS: ELEMENTARY
LIBRARY TEST

GRADE AND SEX	FORM A			FORM B		
	Means*	P.E.	Number of Students	Means	P.E.	Number of Students
<i>5th Grade:</i>						
Total.....	23.94	0.6	106	22.91	0.6	90
Male.....	23.70	0.9	57	22.15	0.9	51
Female.....	24.22	0.9	49	24.00	1.0	39
<i>6th Grade:</i>						
Total.....	28.58	0.5	193	31.37	0.7	131
Male.....	28.57	0.8	89	32.33	1.0	66
Female.....	28.59	0.7	104	30.38	0.8	65
<i>7th Grade:</i>						
Total.....	25.64	0.8	77	23.84	0.9	50
Male.....	25.71	1.2	38	22.04	1.2	25
Female.....	25.56	1.1	39	25.64	1.3	25
<i>8th Grade:</i>						
Total.....	35.81	1.0	53	33.50	0.8	78
Male.....	35.80	1.4	20	32.28	1.1	40
Female.....	35.82	1.4	33	34.79	1.1	38

* Except for 6th grade, figured without the University of Chicago Laboratory School in order to show comparable means for Forms A and B. In other grades only Form A was administered.

.48 while that of men is 6.63. Seniors in high school also show a marked sex difference, but girls vary more than boys. This accentuates the general trend toward higher scores for the girls. That the ranges of the individual scores on the forms are about the same is indicated by the uniformity of the probable errors for each class. This affords some additional evidence of the equivalence of the forms.

Since the mean scores for each level as a whole do not vary markedly and since the reliability coefficients are not greatly

different on the two forms, it is possible to use the tests with some assurance that the differences recorded indicate significant changes in student competence. This permits the use of one form at the beginning of the year or instructional period and of the other at the close.

As an additional check, the section scores and the total scores (minus self-correlation) for one year on each of the levels were correlated, using Form A, for the Freshman year in college, the Sophomore in high school, and the eighth grade in the elementary school. Table 7 shows the coefficients, with their probable errors. In comparing them it should be remembered that the college test contains two sections not included in the high school—Sections I and VII—so that Section II of the college should be compared with Section I of the high school, III with II, and so on. Sections I, II, and III of the high school and of the elementary forms are comparable.

For the high school there is slightly more variation in one school than is found for the group as a whole, but none of the correlations for the total class is over .68 or under .42 in the selected school. The range for the group as a whole is from .36 to .46. The range is not sufficient to justify elimination of any section because it either dominates the test scores or because it is invalid. The scores vary sufficiently from section to section to show that all parts help to differentiate degrees of competence in the use of the library.

The eighth grade was chosen for the elementary test inter-correlations because it was possible to have two large groups typical of different situations—the University of Chicago Laboratory School and the University of Minnesota Training School—in addition to some small groups. Table 7 shows that the greatest variation is for Section III for the total group. That this is the case is not surprising, since a larger portion of the total score was made on the first and second sections than on the third. Consequently, the students who did well on this section would tend to do well on the whole test.

The sex variations noted under the several tables naturally raise the question as to the reliability for the two sexes. An

TABLE 7

INTERCORRELATIONS, SECTION SCORES AND TOTAL (MINUS SELF-CORRELATION)
COLLEGE FRESHMEN, HIGH SCHOOL SOPHOMORES, AND EIGHTH GRADE

SCHOOLS AND SEX	NUMBER OF STUDENTS	SECTIONS*						
		I	II	III	IV	V	VI	VII
Combined colleges:								
Total.....	464	.290 ± .029	.460 ± .025	.540 ± .022	.500 ± .024	.483 ± .024	.341 ± .028	.457 ± .025
Male.....	158	.190 ± .051	.480 ± .041	.530 ± .038	.550 ± .038	.590 ± .035	.401 ± .045	.429 ± .044
Female.....	306	.355 ± .033	.405 ± .032	.523 ± .028	.500 ± .029	.437 ± .031	.300 ± .045	.395 ± .032
Colorado State Teachers College:								
Total.....	204	.335 ± .042	.475 ± .036	.451 ± .038	.475 ± .036	.440 ± .038	.222 ± .045	.410 ± .039
Male.....	76	.351 ± .067	.523 ± .057	.361 ± .067	.400 ± .065	.526 ± .056	.218 ± .074	.317 ± .067
Female.....	128	.389 ± .050	.433 ± .049	.468 ± .047	.409 ± .049	.290 ± .050	.150 ± .059	.405 ± .050
Combined high schools:								
Total.....	194	.360 ± .037	.380 ± .042	.460 ± .038	.460 ± .038	.380 ± .037		
Male.....	86	.330 ± .065	.510 ± .054	.490 ± .055	.430 ± .016	.390 ± .062		
Female.....	108	.410 ± .054	.410 ± .054	.520 ± .047	.510 ± .048	.370 ± .056		
Bloom Township High School:								
Total.....	62	.590 ± .056	.680 ± .047	.420 ± .070	.640 ± .051	.470 ± .067		
Male.....	33	.670 ± .065	.780 ± .046	.304 ± .107	.570 ± .080	.410 ± .098		
Female.....	29	.530 ± .090	.550 ± .088	.470 ± .098	.690 ± .065	.550 ± .088		
Combined Elementary:								
Total.....	124	.319 ± .055	.395 ± .051	.690 ± .032				
Male.....	52	.300 ± .076	.470 ± .082	.380 ± .080				
Female.....	72	.240 ± .075	.334 ± .071	.308 ± .072				

* Sections II-VI of college test and Sections I-V of high school test are comparable. Sections I-III of high school and elementary tests are comparable. Form A of the tests was used.

examination of the coefficients in Table 2 shows that when the colleges as a whole are considered, disregarding class distinctions, the highest for men is .960 and for women .958; while the lowest for men is .888 and for women .776. Table 4 shows that the highest mean score for men is 71.09 and for women 74.83, a difference of 3.74; the lowest for men is 63.06 and for women 64.37, a difference of 1.31. Thus the greatest divergence is seen to be on the high scores.

On the high school level the pattern of the reliability coefficients and mean scores is about the same as for the college. The highest for men is .963 and for women it is .959. The lowest for men is .842 and for women it is .824. The mean scores vary more widely than the college scores, but the greater difference is on the high scores as in college. In both, the higher mean scores are for women, and the higher reliability coefficients for men.

The elementary school shows much less sex variation on the mean scores, the high for boys being 35.80 and for girls 35.82; the low for boys being 22.04 and for girls 24.00. The divergence is much less marked than in the secondary school. The highest correlation coefficient for boys is .87, while that for girls is .91. Table 2 shows this figure for girls in three places. The lowest for boys is .69 and for girls .70. Thus the higher reliability coefficients on this level are for females.

Ruch shows that for students in grades VII–XII there is a marked sex difference on the reliability coefficients for sexes and for the total group on the Stenquist mechanical aptitude tests. For the group as a whole it is .97; for boys, .84; and for girls, .58.¹⁰ This indicates that subject matter is in all probability a major factor in sex differences and that the comparatively slight divergence on the library tests indicates that there is no significant difference between men and women, a conclusion that might be anticipated on logical grounds. The higher scores for the women, with slightly lower reliability coefficients, appear to indicate a somewhat higher ability to utilize the library and greater familiarity with it but slightly less consistency in response. However, there are no marked tendencies that can be

¹⁰ *Ibid.*

attributed generally to sex divergences, although, in some of the sections, as will be shown later, there appear to be areas in which there is a tendency toward sex variations.

Reliability coefficients were also computed for the preliminary test given to the students in the Teachers College at Valley City, North Dakota, during the summer session of 1936. This made it possible to study the differences in the performance of students with and without instruction. Although the groups were rather small in some classes, the results are sufficiently significant to show that formal instruction does have some tendency to stabilize and to raise the reliability of the library test results. This is to be expected on logical grounds. Since standardized achievement tests are based on subject matter that has been accepted as parts of regular courses of study, the fact that the coefficients are as high as they are for groups without instruction indicates that the attempt made in the tests not to penalize the student who could solve the problem of utilizing material without aid, was successful. Table 8 shows the reliability coefficients.

Since changes were made between the preliminary and the revised forms, it is of interest to have a correlation coefficient for the performance of students on the two tests. The group is too small to be considered wholly representative, but the result is significant because of the relative size of the correlation and the probable error. A small group of Freshmen in the University of Chicago High School took the preliminary form at the close of the spring quarter. In the fall they formed a part of the Sophomore class to which Form A of the high school test was administered. The correlation based on the scores of these twenty-one students is $.72 \pm .071$. This is high enough to show that there was a marked tendency for the students who did well on the one test to do well on the other; but the difference is also large enough to suggest that the revision of the tests did have a decided influence on results. The reliability coefficient for Sophomores on Form A, revised test, for this school (based on a split-half count and connected by the Spearman-Brown formula) is $.86 \pm .018$.

THE VALIDITY OF THE LIBRARY TESTS

Although a valid test must be reliable, a reliable test need not be valid. Validity means the extent to which a test measures what it attempts to measure. Various methods of determining

TABLE 8

DIFFERENCES IN RELIABILITY COEFFICIENTS ON THE PRELIMINARY
LIBRARY TEST FOR COLLEGE STUDENTS WITH AND WITHOUT
FORMAL INSTRUCTION IN THE USE OF THE LIBRARY

Class	Correlation Coefficient and Probable Error	Number of Students
<i>Freshmen:</i>		
Total group.....	.888 ± .022	42
With instruction.....	.950 ± .001	18
Without instruction.....	.795 ± .051	24
<i>Sophomores:</i>		
Total group.....	.810 ± .018	169
With instruction.....	.802 ± .022	126
Without instruction.....	.724 ± .049	43
<i>Juniors:</i>		
Total group.....	.840 ± .018	124
With instruction.....	.840 ± .019	104
Without instruction.....	.817 ± .050	20
<i>Seniors:</i>		
Total group.....	.826 ± .027	63
With instruction.....	.890 ± .020	54
Without instruction.....	*	9

* Correlation not made because group was too small.

validity have been established. The terminology varies, but three important concepts are involved.

One of these is the principle of social utility. A second principle emphasizes the judgment of competent persons and is utilized in many ways. It may take the form of specific checking by experts or it may take the form of curricular validation, which is one of the most frequently used of the methods of demonstrating that the test actually does what it claims to do. The third general method of validation is largely statistical, and

includes such procedures as comparison of the returns for groups on different grade levels. To check the validity of the library tests, two general types of evidence were used: the evidence of social utility, and the statistical evidence of consistency by grade.

Three types of procedures were used in establishing the social utility of the tests: validation by curricular content (thus utilizing also the judgment of competent persons), validation by performance in the library, and validation by student estimate of difficulties. The material presented to describe the development of the library tests might be considered as another method of validation, since the methods employed in evolving the test constituted a partial validation of its contents by combining suggestions of librarians on the type of material to be included. The use of a preliminary form to describe student performance at different grade levels is another partial validation.

Validation by curricular content.—Curricular content is presumably based upon social utility; hence conformity to the curriculum is one index of validity. Greene and Jorgensen have shown the assumption to be consistent with modern theories of curriculum construction.¹¹

Library opinion can be cited briefly on the basis of two valuable summaries. One of these is the first *School library yearbook*,¹² issued by the American Library Association. The other is the comprehensive textbook by Ingles and McCague.¹³ The topics suggested in both were checked against the test headings. The minimum courses listed for high schools and for teachers' colleges were considered in dividing the tests for the three levels. Naturally the tests represent a sampling of the fields included, as all tests are necessarily samples of the subject matter tested.

¹¹ *Op. cit.*, p. 133.

¹² American Library Association Education Committee, *School library yearbook*, No. 1 (Chicago: American Library Association, 1927).

¹³ May Ingles and Anna McCague, *Teaching the use of books and libraries* (New York: H. W. Wilson, 1930).

Twenty-six recent courses of study for elementary and high schools, issued by state and city departments of education, were examined, and the grade placement of the topics was studied, both as a means of dividing the tests and as a curricular validation of the content. No attempt was made to analyze a large number of syllabi. A useful description of grade placement of the topics in relation to student performance could be made by a space count of topics in a sufficiently large number of courses and their comparison with student difficulties and successes in the various areas as revealed by answers to questions in the test. Not only were separate library courses examined, but also English and social studies syllabi, for the plan was to find just what topics relevant to use of the library were included, regardless of the way in which they appeared.

The examination of the courses of study showed that each of the topics included in the tests appears in the different grade levels, with emphasis proper for the educational development of the students. "Orientation in the library" is suggested from Grade I through Grade X. The "Meaning of the classification" is included from Grade II through Grade XII. "How to use the catalog" is listed from Grade IV through Grade XII. "Parts of a book," "Use of index," etc., range from Grade II through Grade XII. The "Study of magazines" occurs mostly on the high school level, being listed as Grades VII–XI inclusive, and the "Use of *Readers' guide*" ranges from Grade VII through Grade XII. The "Study of newspapers," though less frequently included, appears first in Grade VII, and last in Grade XII. "Study of encyclopedias" is listed as low as Grade IV and as high as Grade XII. "Dictionaries" are introduced first in Grade III and continue through Grade XII. "Special indexes" appear less frequently, but range from Grade VI through Grade XII. "Reference books" are introduced in Grade V and continue through Grade XII.

This evidence shows that the tests are validated from the point of curricular content. Since experts in the field of curriculum construction have selected these items as beneficial to the educational development of students, this tabulation shows

also that the items of the tests are justified on the basis of social utility.

Validation by performance in the library.—The evidence provided by the performance of students was gathered by E. Pearl Hess, librarian of the Bloom Township High School, through questions looked up by selected groups of students during regular class periods, their activities being recorded by Miss Hess. In deciding on the pupils to be used in the experiment, two or three boys and the same number of girls ranking highest and lowest in each form of the test in both the Senior and Junior years were chosen.

Eleven questions were selected from a list that had been kept of difficulties encountered in regular work and suggestions by the head of the English department. The attempt was made to choose a group of problems the answers to which could be found in a class period. They were listed in the order of importance for performance validation, so that the least loss would be felt if the student was unable to complete the work during a regular class period.

Two students were taken at a time, the librarian working with them as she would ordinarily work with students coming to the library to look up material in connection with courses. In every way the attempt was made to create the atmosphere of regular work rather than that of a test situation. The students had been very much interested in the tests, and desirous of doing work in the library. Since no grades were given out before the work in the library, the students did not know whether they were chosen because of high or low grades on the test. In the few cases where there was some evidence of emotional strain this was relieved by the librarian before the work began.

The questions included three on the catalog, checking on the use of title entries, abbreviations, and the filing rule of omission of the initial article. The use of the catalog and general reference books was required in one calling for the listing of material on the life of Field Marshal Haig. The use of special indexes was necessary for the finding of a poem known by title. Use of year-books, dictionaries, and special reference books in the fields of

art, history, and literature was required in the remainder of the eleven questions.

The librarian recorded the activities of the students and gave aid only in analyzing the questions, according to her usual procedure of informal instruction. She noted particularly evidence bearing on eight questions that had been listed in advance.

1. Where does the student go first? Does he stop to analyze the question? Or does he always go to a favorite reference book regardless of the type of question?

2. Under what headings does he look in trying to locate the information? Does he follow up cross-references? Does he make use of indexes, tables of contents, etc.?

3. Does he use the most specific book on the subject or only general reference books?

4. Can he do his work independently, or does he have to stop and ask for help? If the latter, what type of questions does he ask, and what kind of help is given?

5. Is he satisfied with a brief and perhaps inadequate answer to the question, or does he carry on and find more material in other sources?

6. Does he always make the same kind of error with all questions, or is this variable?

7. How long does it take him to find the material? How many trials does this involve?

8. Does he show originality in handling assignments? Or does he follow only established routines?

This material has been presented in detail as case studies¹⁴ and includes the results of the retesting that followed brief instruction given at the close of the observation of student behavior in the library. The results of the performance validation is summarized in the following conclusions, and in Tables 9-14 inclusive. Much of the material presented in the case studies has not been summarized in tabular form because of the danger of distorting the data and of attributing to a small number of cases too much statistical significance. However, some of the test results as related to students' behavior in the library could be tabulated so as to make the relationships plainer and to emphasize the general conclusions drawn from the facts. That the performance of individual students does not always provide a

¹⁴ For case studies see author's unpublished Doctor's thesis, *op. cit.*

TABLE 9

RELATIONSHIP BETWEEN TEST SCORES AND PERFORMANCE, BLOOM
TOWNSHIP HIGH SCHOOL: AMOUNT OF ANALYSIS AID REQUIRED
RANKED GROUPS OF SENIORS

SENIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM A	TOTAL NUMBER OF QUESTIONS ON WHICH AID WAS REQUIRED											SENIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM B	TOTAL NUMBER OF QUESTIONS ON WHICH AID WAS REQUIRED										
	1	2	3	4	5	6	7	8	9	10	11		1	2	3	4	5	6	7	8	9	10	11
1*.....	×											1*		×									
2.....					×							2		×									
3.....				×								3			×								
4.....						×						4				×							
5.....		×										5			×								
6.....	×											6		×									
7.....				×								7						×					
8.....					×							8			×								
9.....						×						9			×								
10.....						×						10				×							
11.....				×																			
12.....		×																					

* Nos. 1-6 scored above the class mean, the others below the class mean.

TABLE 10

RELATIONSHIP BETWEEN TEST SCORES AND PERFORMANCE, BLOOM TOWNSHIP
HIGH SCHOOL: AMOUNT OF ANALYSIS AID REQUIRED
RANKED GROUPS OF JUNIORS

JUNIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM A	TOTAL NUMBER OF QUESTIONS ON WHICH AID WAS REQUIRED											JUNIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM B	TOTAL NUMBER OF QUESTIONS ON WHICH AID WAS REQUIRED										
	1	2	3	4	5	6	7	8	9	10	11		1	2	3	4	5	6	7	8	9	10	11
1*.....					×							1†			×								
2.....	×											2	×										
3.....												3	×										
4.....				×								4				×							
5.....				×								5			×								
6.....		×										6			**								
7.....						×						7						×					
8.....						×						8									×		
9.....						×																	
10.....					†																		

* Nos. 1-6 scored above the class mean, the others below the class mean.

† Nos. 1-5 scored above the class mean, the others below the class mean.

‡ Required aid on 1 and omitted 4.

** Required aid on 2 and omitted 1.

definite parallel with test results does not invalidate the use of the tests for diagnostic purposes. The amount of individual

TABLE 11

RELATIONSHIP BETWEEN TEST SCORES AND PERFORMANCE, BLOOM TOWNSHIP HIGH SCHOOL: TYPES OF QUESTIONS ON WHICH AID WAS REQUIRED BY STUDENTS SCORING ABOVE AND BELOW CLASS MEANS ON LIBRARY TEST

TYPE OF QUESTION	NUMBER OF STUDENTS REQUIRING AID	
	Scored Above Class Mean Library Test	Scored Below Class Mean Library Test
<i>Catalog entries:</i>		
Title and abbreviations		
Question 1.....	0	1
Question 9.....	0	0
Title: Initial article		
Question 11.....	0	1
<i>Dictionaries:</i>		
Question 7.....	0	0
Question 8.....	8	13
<i>Yearbooks:</i>		
Question 4.....	9	15*
<i>Special indexes and collections:</i>		
Question 3.....	18	15*
<i>Special reference books:</i>		
Art		
Question 5.....	15	13*
History		
Question 6.....	9	9†
Literature		
Question 10.....	9	11
Number of students.....	23	17

* Does not include one student who did not attempt question.

† Does not include two students who did not attempt question.

agreement is very large for so few cases, and the group summaries substantiate the trends shown by the test.

The conclusions indicate that, in general, performance parallels test results. They may be summarized as follows:

1. Students scoring high in the test on the whole tended to work more quickly and efficiently in the library. This is shown by the fact that the average time for students who had scored above the class mean on the test was forty-one minutes and

TABLE 12

RELATIONSHIP BETWEEN TEST SCORES AND PERFORMANCE, BLOOM TOWNSHIP HIGH SCHOOL: NUMBER OF QUESTIONS ON WHICH AID WAS REQUIRED BY STUDENTS SCORING ABOVE AND BELOW CLASS MEANS ON LIBRARY TEST

NUMBER OF QUESTIONS ON WHICH AID WAS REQUIRED	NUMBER OF STUDENTS REQUIRING AID	
	Students Scoring Above Class Mean Library Test	Students Scoring Below Class Mean Library Test
0.....		
1.....	5	1
2.....	4	2
3.....	5	2
4.....	6	3
5.....	2	1
6.....	1	7
7.....	0	0
8.....	0	0
9.....	0	1
10.....	0	0
11.....	0	0
Total students.....	23	17
Total questions requiring aid. . .	68	79
Mean questions requiring aid. . .	2.96	4.65

forty-four seconds for performance, while that of students who had scored below the class mean was forty-nine minutes and fourteen seconds. Hence, lower scores on the test and omission of items apparently are directly connected with student ability and grasp of subject, indicating that the test is not too long.

2. In general, the same type of difficulties was revealed by performance as by the test. In a few cases students showed more ability in using the actual catalog than in interpreting the

entries, but sampling was much less complete in performance than in the test scores. Consequently, the test apparently is satisfactory in its diagnosis of difficulties.

3. Use of special reference books indicated the same indefiniteness of knowledge shown by the test. Particularly, a lack of knowledge of special indexes was revealed. The librarian explained this in part by the reference policies in use in the li-

TABLE 13
RELATIONSHIP BETWEEN TEST SCORES AND PERFORMANCE, BLOOM
TOWNSHIP HIGH SCHOOL: NUMBER OF ERRORS IN USE OF
CATALOG: RANKED GROUPS OF STUDENTS

SENIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM A	NUMBER OF ERRORS USE OF CATALOG				SENIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM B	NUMBER OF ERRORS USE OF CATALOG				JUNIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM A	NUMBER OF ERRORS USE OF CATALOG				JUNIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM B	NUMBER OF ERRORS USE OF CATALOG			
	0	1	2	3		0	1	2	3		0	1	2	3		0	1	2	3
1*.....	X				1*	X				1*	X				1†	X			
2.....		X			2			X		2		X			2	X			
3.....	X				3	X				3	X				3	X			
4.....	X				4	X				4	X				4		X		
5.....	X				5	X				5	X				5		X		
6.....	X				6	X				6		X			6	X			
7.....	X				7	X				7			X		7		X		
8.....				X	8			X		8	X				8			X	
9.....		X			9		X			9		X							
10.....		X			10			X		10		X							
11.....		X																	
12.....		X																	

* Nos. 1-6 scored above the class mean, the others below the class mean.

† Nos. 1-5 scored above the class mean, the others below the class mean.

braries, where students selected reading material directly from special typed bibliographies rather than by the use of printed indexes. Books are arranged in special sections such as essays, short stories, etc., owing to the necessity for handling large groups of students every hour in the library. Table 11 shows the types of questions on which aid was required, and the amount needed on special indexes is apparent.

4. Students all needed assistance in analysis of reference questions and their association with special reference books.

But in general, the students with low test scores required much more aid than those with higher scores, indicating that the test does differentiate the students more able to work independently with library resources from those who are less able. Tables 9 and 10 show the relationship between the amount of analysis aid required and the ranking of students on the test. Table 12 summarizes the relationship between the number of questions

TABLE 14
RELATIONSHIP BETWEEN TEST SCORES AND PERFORMANCE, BLOOM
TOWNSHIP HIGH SCHOOL: USE OF SPECIAL SECTION OF
DICTIONARY: RANKED GROUPS OF STUDENTS

SENIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM A	USED SPECIAL SECTION IMMEDIATELY		SENIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM B	USED SPECIAL SECTION IMMEDIATELY		JUNIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM A	USED SPECIAL SECTION IMMEDIATELY		JUNIORS ARRANGED IN RANK ORDER BY TEST SCORES FORM B	USED SPECIAL SECTION IMMEDIATELY	
	Yes	No		Yes	No		Yes	No		Yes	No
1*	×		1*	×		1*	×		1†		×
2		×	2	×		2	×		2	×	
3		×	3	×		3		×	3	×	
4		×	4	×		4		×	4		×
5	×		5		×	5	×		5		×
6	×		6	×		6		×	6		×
7	×		7		×	7		×	7		×
8		×	8		×	8		×	8		×
9		×	9	×		9		×			
10		×	10		×	10		×			
11		×									
12		×									

* Nos. 1-6 scored above the class mean, the others below the class mean.
† Nos. 1-5 scored above the class mean, the others below the class mean.

on which aid was needed and the grouping of students as scoring above or below the class mean on the library test. This shows that those above the mean on the average required aid on 2.96 questions in the performance problems, while those below the class mean required assistance on 4.65 questions. Table 14 shows specifically that the degree of specific knowledge of the characteristics of reference books such as the biographical section of the unabridged dictionary is closely related to the weaknesses revealed by the test.

5. There is some indication that difficulties with alphabetical arrangement had a bearing on problems of students in the lower group of scores.

6. There is some difficulty in the use of specific reference books that may be attributed to special connotation of book titles. This appeared to be the case with Reinach's *Apollo*.

7. Retesting after brief instruction tended to prove the validity of the tests. Although the period of instruction was too brief and the group retested too small for the results to be definitive, they are highly suggestive. The better students appear to acquire knowledge of library resources for themselves, but have weak spots that instruction aids. The poor students show general weakness, with good response to instruction. In general, the changes for poor students and the stability of grades for good students indicate that the tests possess validity and can be used as measures of the differences in ability to utilize library resources.

Validation by student estimate of difficulties.—The material for the validation of the tests in terms of their social utility to college students was supplied by Lovell H. Strauss, librarian of George Williams College. Groups of students were selected for testing, and after tabulation of scores a few from each year were selected for interviews. In choosing the group for further investigation, some scoring high and some scoring low on each of the test forms were listed. Mr. Strauss worked out an interview check list on which he also wrote general impressions of the students. The information gained by this method was summarized and conclusions drawn. The check list prepared by Mr. Strauss was a very comprehensive inventory, covering the scholastic, social, and literary background of students, as well as the student's evaluation of the difficulties encountered in using the library.

The students selected showed sufficient variety in their personal and scholastic backgrounds to indicate that the test is successful in differentiating students on the basis of ability to utilize independently and successfully library resources and that the general areas of difficulty are shown by the test re-

sults. Even though special reasons such as language difficulty may account for part of the failures, the tests can be used for diagnosing general types of trouble and for planning remedial

TABLE 15
RELATIONSHIP BETWEEN TEST RESULTS AND STUDENT ESTIMATES
OF DIFFICULTY: USE OF CATALOG

STUDENTS ARRANGED IN RANK ORDER BY TEST SCORES	CHIEF DIFFICULTIES SHOWN BY TEST			GENERAL ABILITY IN USE OF CATALOG SHOWN BY TEST		STUDENT ESTIMATE BY INTERVIEW			
						Used Catalog Before Going to Shelves		Felt Use of Catalog Difficult	
	Author and Title Entries	Subject Entries	Analyt- ics and Special Card	Good	Poor	Habit- ually	Seldom	Yes	No
<i>High group:*</i>									
1.....		×		×		×			×
2.....				×			×		×
3.....		×	×	×			×		×
4.....	×		×	×		×			×
5.....	×		×	×			×		×
6.....	×			×			×		×
7.....			×	×			×		×
8.....	×		×	×		×			×
<i>Low group:†</i>									
9.....	×		×	×			×		×
10.....				×			×		×
11.....					×		×		×
12.....	×			×			×		×
13.....	×		×		×		×	×	
14.....					×		×	×	
15.....					×		×		×
16.....	×	×			×		×	×	

* Scored above the class mean.
† Scored below the class mean.

work. It also appeared that the students standing high on the test as a rule had had more opportunity to use libraries. But this was not the only determining factor, for some students with less favorable social and economic backgrounds have achieved success when given help in college, and others who have had opportunities have not used them. In these cases the general

idea seems to be "let the librarian do the work." While the librarian is perfectly willing to assist students, this attitude does not help students to learn to help themselves. Some of the relationships between test results and student estimate of difficulties appear in Tables 15, 16, and 17.

TABLE 16
RELATIONSHIP BETWEEN TEST RESULTS AND STUDENT ESTIMATES
OF DIFFICULTY: USE OF INDEXES

STUDENTS ARRANGED IN RANK ORDER BY TEST SCORES	CHIEF DIFFICULTIES SHOWN BY TEST				GENERAL ABILITY SHOWN BY TEST		INTERVIEW STATEMENT: USED INDEXES FREQUENTLY							
							Readers' Guide		Special Indexes		Newspaper Index		Gov't. Docs. Index	
	Read- ers' Guide	Spe- cial In- dexes	News- paper In- dex	Gov't Docs. In- dex	Good	Poor	Yes	No	Yes	No	Yes	No	Yes	No
High group:*														
1.....		X			X		X		X		X			X
2.....		X		X	X		X			X	X			X
3.....		X			X			X		X	X			X
4.....		X			X		X			X	X			X
5.....		X		X	X			X		X	X			X
6.....	X	X		X		X		X		X	X			X
7.....	X	X				X		X		X	X			X
8.....	X	X		X		X		X		X	X			X
Low group:†														
9.....	X	X	X	X		X		X		X	X			X
10.....	X	X	X	X		X		X		X	X			X
11.....	X	X	X	X		X		X		X	X			X
12.....	X	X	X	X		X		X		X	X			X
13.....	X	X	X	X		X		X		X	X			X
14.....	X	X	X	X		X		X		X	X			X
15.....	X	X	X	X		X		X		X	X			X
16.....	X	X	X	X		X		X		X	X			X

* Scored above the class mean

† Scored below the class mean.

Statistical evidence of validity.—Various statistical measures of validity are well established. In general they involve correlation with other measures of assumed validity, such as school marks and other presumably valid measures of achievement.

Since there are no thoroughly standardized library tests of

demonstrated validity, they are not available as a standard. Correlation of the library tests with school marks and achievements might be significant. To make such correlations one should take selected groups, trained by specific courses of study on the use of the library, and compare test grades made by such students with others not so trained. There are two objections

TABLE 17
RELATIONSHIP BETWEEN TEST RESULTS AND STUDENT ESTIMATES
OF DIFFICULTY: USE OF SPECIAL REFERENCE BOOKS

STUDENTS ARRANGED IN RANK ORDER BY TEST SCORE	PERFORMANCE ON TEST			INTERVIEW STATEMENT: FAMILIAR WITH SPECIAL REFERENCE BOOKS			INTERVIEW STATEMENT: FELT DIFFICULTY IN USE OF SPECIAL REFERENCE BOOKS	
	Strong	Fair	Weak	Many	Some	Few	Yes	No
<i>High group:*</i>								
1		/		/				/
2		X			/		/	
3		/			/			/
4			X			/		/
5		X		/				/
6		/			/		/	
7		X				X	X	
8			X		X		X	
<i>Low group:†</i>								
9			/			/	/	
10			X			X	X	
11			X			X	X	
12			X		X		X	
13		X		X			X	
14			X			X	X	
15			X			X	X	
16			X			X	X	

* Scored above the class mean.

† Scored below the class mean.

to this method of validation: first, it is hard to tell what constitutes a relevant course of study; second, it is hard to determine to what extent the same information may be gained by informal rather than formal methods of instruction. The library tests could be used to supply data relevant to the further study of the two problems indicated.

The students with some formal instruction performed more consistently on the preliminary test than did the others. The reliability coefficients for the two groups supply further evidence of test validity. The performance on the revised test of the students who had been given more instruction than any of the others indicates the same thing. The reliability coefficients for these Freshman groups were $.922 \pm .019$ on Form A and $.966 \pm .009$ on Form B, each being the highest coefficient obtained by any groups on either form.

If we could assume that some of the qualities tested by other standard tests were also tested by the library tests, validation by correlation with other achievement tests such as reading, English, social studies, or psychological tests would be highly relevant. Such an assumption would be open to very grave doubts and should not form the basis for validation until further investigation presents evidence tending to show the relation of library skills to other skills. The exact nature of library skills must be established before the assumption of similarity with other skills can be accepted as a basis for validation.

Comparison of the test grades from the various schools used in the investigation establishes validity on the grounds of returns from widely spaced groups and of rise in percentage of success. In the colleges the highest scores for Freshmen were made at Fenn, which reported the largest amount of instruction and used small picked groups. Colorado, which gave some library instruction to at least part of the students, stands next in mean scores. The groups used for upper classes were too small for the percentages to be entirely significant, but in general there is a 10 per cent difference in the test performance of students in junior and in senior college, while for individual schools this difference is about 7 per cent. About the same amount of difference is found on the high school level. The individual schools show about 10 per cent difference between the scores of the first two and the last two years, while the total group shows a difference of about 11 or 12 per cent. The individual schools on the elementary level show less difference from grade to grade,

but the combined schools show a difference of about 12 per cent between the fifth and the eighth grades.

One other evidence of a statistical nature was available in the standing of ranked groups. It is based on the assumption that groups differing markedly in ability in English would tend as a whole to vary in the same way on the library tests. The English instructor and librarian at St. Patrick's High School, Mr. John B. Leahy, was asked to rank the Freshmen and Senior students according to their ability as shown on English tests, ranking them high, medium, and low. Then their library test grades

TABLE 18
TOTAL MEAN SCORES, RANKED GROUPS, HIGH SCHOOL LIBRARY
TEST: ST. PATRICK'S HIGH SCHOOL (MALE)
FRESHMEN AND SENIORS*

RANKED GROUPS	FRESHMEN		SENIORS	
	Form A	Form B	Form A	Form B
Total.....	41.92	40.22	50.58	55.64
High.....	53.47	52.43	57.88	66.06
Medium.....	40.18	42.71	50.17	51.62
Low.....	30.64	25.86	37.89	40.17

* Ranked by English teacher on basis of English tests.

were averaged for the classes as a whole and for each group. Table 18 shows that mean scores on the library tests follow the same order as the ranking of the groups by the English tests.

Thus, from the point of social utility and from statistical evidence the tests are sufficiently valid and reliable to be used in the study of differences in the ability of students to use libraries. The evidence on the basis of social utility was presented in three different ways: first, by a sampling of courses of study, thus giving curricular validation; second, by performance of high school students in the library; and third, by student estimate of difficulties on the college level. Statistical evidence is in terms of returns from widely spaced groups and rise in percentage of success, with some discussion of reliability co-efficients, and other possible methods of validation.

DIFFERENCES IN THE ABILITY TO USE THE LIBRARY

The differences in the ability to use the library that are described in the following pages are based on the scores for the whole test and for the sections and subsections. Differences for which the tabulation of separate items would be necessary are not described. Further research is under way.

The sectional scores show a general pattern for each of the three educational levels. The elementary tests follow a single pattern throughout, the order being Sections I, II, and III, with no sex variations. The high school tests follow the general order of I, II, IV, V, III, but with variations, particularly in the last two, and with some sex variations. These are shown in detail in Table 20. The general order for colleges is II, V, III, I, VI, IV, and VII, but with several variations, especially in the Sophomore year. In all tests, the section on the finding of the individual book ranks highest in each of the levels, since Section I in elementary and high school and Section II in the college test are devoted to this topic. Elementary and high school forms place next the section on the use of books, which includes parts of books, dictionaries, encyclopedias, and special reference books in various fields. The comparable section in college (III) comes third, being surpassed by the section on newspapers and newspaper indexes which comes third in the high school. There is no section on newspapers in the elementary test. Of the remaining two sections in the high school test, evaluation of authenticity of material precedes the use of the magazine index. This is true in the college tests, Section VI coming before IV, but Section I, on differentiation between services of library departments, precedes the authenticity section. The one in the college test that stands lowest in rank order is that devoted to the government document indexes.

Table 19 shows the exact order of the sections for each class and sex on the college test. The sex variations in the college tests are most frequently connected with the order of Sections I and III. In the Freshman, Sophomore, and Junior years the males tended to do better on III than I on Form A, and in the Senior year the females tended to do better on I than III on the

TABLE 18

COLUMBIAN LIBRARY TEST SECTIONS ARRANGED IN ORDER OF MEAN SCORES, BY CLASS AND SEX

SUBSECTIONS				SCORES			
CLASS A		CLASS B		CLASS A		CLASS B	
♂*	Me*	♀	Me	♂	Me	♀	Me
II	II	II	II	II	II	II	II
V	V	V	V	V	V	V	V
I	I	III	III	I	I	III	III
III	III	I	I	III	III	I	I
VI	VI	VI	VI	IV	IV	VI	VI
IV	IV	IV	IV	IV	IV	IV	IV
VII	VII	IV	IV	VII	VII	VII	VII

* T = Total, M = Mean, ♀ = Female.

TABLE 20
HIGH SCHOOL LIBRARY TEST SECTIONS ARRANGED IN ORDER OF MEAN SCORES, BY CLASS AND SEX

FRESHMEN						SOPHOMORES						JUNIORS						SENIORS					
FORM A			FORM B			FORM A			FORM B			FORM A			FORM B			FORM A			FORM B		
T*	M*	F*	T	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T	M	F
I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II	II
IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III	III
V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

* T = Total; M = Male; F = Female.

form. In general, there is little sex variation in the order of sections for the high school. These are practically all connected with the relative placement of Sections II and IV, and of Sections III and V. The mean scores, in terms of percentages of the total possible scores, for the main sections are given in Table 21.

Locating specific books.—There is a definite progression in the ability to use the card catalog for author and title entries as shown by the sectional mean scores of the tests when made directly comparable as percentages of the total possible score. This was tested by Section I, subsection A of the elementary and high school tests, and by Section II, subsection A of the college tests. The distinction between high school and college is sharper than between elementary and college. The explanation for this is probably the difference in organization and size of collections, making dependence on the catalog more essential in colleges.

Locating specific subjects.—Subdivision B in the same sections of the tests takes up the subject entries in the catalog. As in author and title entries, there is some general gain from level to level in the interpretation and use of subject entries. The variations are wide enough to justify item tabulations to ascertain specific differences. Types of problems include the use of personal names, technical decisions involving use of period headings in history, and specific kinds of entry words.

Interpreting special types of catalog cards.—Other divisions of Section I of the elementary and high school tests, and of Section II of the college tests, are devoted to the special types of catalog cards such as editors, joint authors, and illustrators. The results are not markedly different from those of the kinds of entries already discussed. The differences between levels are less sharply drawn and there are some places where an orderly progression is not shown. The variations in the elementary grades might well be expected on the ground that students had little occasion to use this type of entry. In general there is a rise in performance from the eighth grade through high school, but with some variations. Divergences in college are more marked than in secondary school.

TABLE 21
SECTIONAL SCORES ON LIBRARY TESTS*

CLASS	SECTIONS†						
	I	II	III	IV	V	VI	VII
<i>College test:</i>							
Freshmen							
Form A.....	39.00	52.83	38.66	19.07	51.50	21.20	18.00
Form B.....	33.40	51.00	42.00	15.07	48.50	30.60	19.20
Sophomores							
Form A.....	40.30	55.40	42.51	23.13	52.80	23.40	14.10
Form B.....	30.70	52.98	44.17	18.80	51.70	33.20	18.20
Juniors							
Form A.....	44.10	56.23	42.51	31.20	48.75	22.00	12.90
Form B.....	31.80	52.40	43.29	19.73	48.90	25.40	17.10
Seniors							
Form A.....	42.10	56.70	44.23	33.33	55.95	21.60	13.70
Form B.....	36.00	56.85	43.11	19.87	47.55	26.00	13.80
<i>High school test: ‡</i>							
Freshmen							
Form A.....	46.63	26.49	13.13	22.35	4.20		
Form B.....	47.30	24.97	13.67	19.70	6.40		
Sophomores							
Form A.....	51.15	33.97	19.07	43.25	14.20		
Form B.....	51.15	36.17	10.93	35.90	15.40		
Juniors							
Form A.....	50.88	36.49	23.00	42.85	11.80		
Form B.....	52.48	36.57	16.33	38.20	18.20		
Seniors							
Form A.....	52.75	36.60	30.60	31.65	7.20		
Form B.....	54.73	39.66	34.67	40.20	18.20		
<i>Elementary test: ‡</i>							
Fifth grade							
Form A.....	35.00	11.64	0.27				
Form B.....	34.17	9.36	0.47				
Sixth grade							
Form A.....	39.65	18.00	1.93				
Form B.....	46.15	13.96	1.27				
Seventh grade							
Form A.....	33.87	19.80	2.40				
Form B.....	33.03	15.28	1.33				
Eighth grade							
Form A.....	41.60	36.32	11.80				
Form B.....	41.15	31.08	6.93				

* Mean scores in terms of percentages of total possible score.

† Sections I-III of high school and elementary tests are comparable. Sections I-V of high school and II-VI of college tests are comparable.

‡ Mean scores for high school and elementary tests are figured without returns from the University of Chicago Laboratory School, except for the sixth grade, as in all other grades only Form A of the test was administered.

Associating reference topics with classification numbers.—There is a gain in each of the educational levels in the ability to associate the reference topics with the classification. The greatest change comes in the eighth grade and in high school, the Seniors in high school being far above college students. That the high school Seniors surpass all the college students might be surprising were it not for the fact that open-shelf collections in high school are the common practice, while the college student is often called on to use the subject catalog more frequently because of closed stacks. With only five exceptions, the sex distinction is clearly in favor of males.

An item tabulation would bring out many additional facts and probably establish numerous differences both on the grounds of grade and of sex. Differences in the various subject fields doubtless would vary extensively. Such questions as the relative ability of Seniors and Freshmen to find topics in history, art, and science would be determined by the study of the individual items of the section. As in the other types of entries, the association of the results with definite policies of instruction would help to clear up the problem and to show the differences due to instructional methods.

Using parts of books.—There appears to be a definite progress in the understanding of the parts of books and the type of information which can be secured by the use of such aids as tables of contents, indexes, and title-pages. No significant divergences appear on either of the forms in the answers to the first division of Section II of the elementary and high school tests and Section III of the college test. There is some evidence of a sex difference in favor of women. Teaching methods and grade placement of the lessons on this subject might well be considered in connection with detailed analysis of the responses to the individual questions in this section.

Using special indexes.—General weakness in the knowledge of special indexes to poetry, recitations, short stories, essays, and similar material is indicated on all levels by the very low scores. The highest for the elementary level amounts to only about 13 per cent of the total possible score, and this is the exceptional performance by the eighth grade on Form B. The

next highest is 6 per cent in the seventh grade on the same form. In high school the highest is the Senior year, with about 11 per cent, also on Form B. On the same form the Sophomores and Juniors score about 10 per cent. All others are much lower, and on Form A none is over 2 per cent. In college the highest is 16 per cent on Form B, Sophomores. The scores are all low enough to indicate that neither men nor women have learned to utilize the special indexes that supplement the card catalog as a means of finding specific material. Additional study based on item tabulation is necessary to reveal whether poetry indexes are better known than essay or fiction lists, whether students distinguish between indexes and collections, and whether variations in local collections are important factors in differences. In this connection a check on instructional methods would be important, since it was shown in the performance work at Bloom Township High School that the use of special typed lists tended to lessen the dependence on printed aids for finding poems, etc.

Differentiating between various encyclopedias.—There is a tendency to be more discriminating in the use of encyclopedias during the Sophomore and Junior years of high school than in the Freshman and Senior years, and in the Freshman and Sophomore years in college than in the Junior and Senior years. This may be because of the greater emphasis on special material during the Senior year in high school and in the Junior and Senior years of college. To determine this it would be necessary to study the amount of use of encyclopedias by the various classes and to compare them with other reference material used. An item tabulation of the questions would reveal the exact variations, according to the different types of questions included. The fact that the general level of discrimination is higher during college than during high school, except for rather minor variations, can be accepted on the logical grounds of maturity of viewpoint. No decided sex trend is apparent, for the variations from year to year seem to follow no definite pattern.

Using unabridged dictionaries.—The evidence from the two forms for the college and high school levels is rather contradictory, but in general indicates progress with age in the ability to

use effectively unabridged dictionaries. Form A shows more orderly progression than Form B in the college. There is about equal variation on the two forms for high school. Further study is necessary to determine the relationship between secondary and elementary achievement. This would have to be based on an item tabulation, since there is no separate section devoted to this topic in the elementary test. Because the use of unabridged dictionaries is usually introduced in the junior high school, some facts of importance to the study of instructional methods might be determined by such a tabulation. In addition, a check by individual questions would reveal the differences on all levels in the use of special features, such as synonyms, keys to pronunciation, illustrations, and divided pages, at the various grade levels.

Associating types of reference questions with specific reference books.—There seems to be a general weakness in ability to associate types of questions with specific reference books to judge by the behavior of students in high school, and by interviews with college students, for the mean scores on this section of the test are low. Ability generally progresses from the Freshman year through college. Item tabulation would show which particular fields most influence the results—whether history, art, or other specific topics—and what books are best and least known. This should also be supplemented by a check of the holdings, although the books selected appear in standard reference guides and, therefore, the chances that the libraries hold the books are above average. Additional study in connection with courses of instruction and instructional methods would be important, particularly as it was shown at Bloom Township High School that certain local practices, such as the preparation of special bibliographies, apparently influence the amount of familiarity with special reference books acquired by the students and that there was a variation in actual titles known because of specific courses taken by the students.

Utilizing magazines and newspapers.—The best understanding of the use of the *Readers' guide* appears to be in the Senior year of high school, and in senior college, though the second

trend is not equally pronounced in both forms. The first place that shows a marked difference is between the seventh and eighth grades. The high score attained in the Senior year in high school is probably due to the emphasis on magazines in grades 9, 10, and 11. This was evident from the checking of courses of study from state and city education departments and, therefore, it is not surprising to find superior achievement in the Senior year.

The reason why college Freshmen do less well is not clear. The explanation may lie in the fact that a great many of the Freshmen included in the college testing came from smaller schools and may not have had the opportunity to use the *Readers' guide*, afforded to students in the more metropolitan communities where the high school testing was done. Part of the showing of college students on the *Readers' guide* in contrast to their behavior on the *New York times index* may be due to the difference in format between the two exercises. However, the fact that the difference is less marked in the high school tends to disprove this hypothesis and to place the explanation more squarely on the point of greater familiarity with the index by the high school students used in the testing.

The fact that college students show more familiarity with the newspapers themselves than do the high school students seems to substantiate the theory that the degree to which students had probably been exposed to the use of the *Readers' guide* played a large part in the test results. This familiarity was to be expected on purely logical grounds. There are some sex variations, usually in favor of women. Here again an item tabulation would reveal such points as differences in the use of special sections of newspapers, interpretation of the characteristics of newspaper stories, particularly in relation to the use of material for reference purposes. In connection with the study of the items, the particular emphasis on magazines and newspapers in the courses used in the schools would be most pertinent.

In general it appears that there is a direct relationship, especially on the college level, between knowledge of newspapers and interpretation of indexes. The low magazine-index scores

may be explained by the infrequent use of indexes in secondary schools, which handicaps college Freshmen from the smaller communities.

Using indexes to government documents.—The closing section of the college test, devoted to the government-document indexes, raises the question whether performance on these indexes corresponds to that shown on the *Readers' guide* and the *New York times index*. The Freshmen are about as able to interpret government indexes, which they probably have not used, as the *Readers' guide*. This is not true of the other classes in general, though it does apply to the Sophomore year, as indicated by the results on Form B.

An item tabulation would bring out the chief sources of difficulty, such as authors, special symbols, or technical details. Inspection without tabulation suggests that the technical features are the chief difficulty; if so, the section differentiates the student who really knows the material from the one who does not. It would be useful to study the returns on the document indexes in relation to specific college classes, and to check on the type of assigned readings, to see how far instructional differences influence the results.

Discriminating between services of library departments, and evaluating authenticity of material.—The two topics may be considered together, though only the college test contains a section on the differentiation of departmental services. Both sections emphasize the quality of evaluation, as against the knowledge of tools, as shown by the other sections of the test. In general, there is a growth in discrimination of services rendered by the various departments of a library from Freshman through Senior years, with variations mainly attributable to sex. But there is no definite sex pattern. An item tabulation would help the interpretation.

There does not seem to be a parallel between the sections on the differentiation of services to departments and the evaluation of authenticity of material, although both call for discrimination rather than knowledge of tools. An item tabulation would reveal the differentiation between types of questions and

might give a definite curricular connotation to the differences in behavior of students, particularly in high schools. One of the schools included in the testing so far surpassed the others that the head of the English department was asked for suggestions as to possible explanations. It was stated that in certain assignments the students had been given opportunities of differentiating between sources as to their relevancy and authenticity.

RELATIONSHIP BETWEEN LIBRARY SKILLS
AND OTHER SKILLS

Although the sampling was too limited to establish definitely any specific relationship between library and other skills, the correlations made between the performance of students on the library tests and some intelligence and subject-matter tests are suggestive. These tend to establish the fact that intelligence and knowledge of school subjects make for competence in library use. This should give librarians no great surprise. General reading tests conform more closely than do the other subject-matter tests to the pattern of the intelligence tests. Social science, which like reading and English depends largely on the library as its laboratory, shows less wide variation than English.

There is much less tendency for close relationship between library test results and Stanford achievement tests in the junior high school than there is between library and subject tests in college. This is probably owing to the complexity of this battery of tests, and emphasizes the need for correlating library test results by sections with very specific skill tests in the subject fields. This is shown clearly by the comparison of the coefficients obtained for the library tests and other measures and those between subject and intelligence tests. The studies of Wrightstone¹⁵ and Ruch and Stoddard¹⁶ indicate that there is a general trend toward coefficients up to .55 between the various subject tests, probably owing to the effect of integrated curriculums, but that the differences are sufficiently important to

¹⁵ J. W. Wrightstone, "Correlations among tests of high-school subjects," *School review*, XLIII (1935), 199-201.

¹⁶ *Op. cit.*, pp. 218-19.

indicate that the skills tested are appreciably different and that the various tests add to the general knowledge of student competence in various lines.

SUMMARY

The investigation of student competence in the use of the library has attempted to show, first of all, that the tests used in the study were sufficiently reliable and valid to provide the means for more detailed investigation of the topics on the basis of test-item tabulations. The methods employed in the construction of the tests helped to validate them, through the use of opinions of librarians and through the returns from a preliminary test showing actual student performance. The other methods of validation were based largely on the principle of social utility—including curricular validation, performance of high school students in the library, and interviews with college students to ascertain their estimate of difficulties in relation to test results. The other measures were statistical, including the increases in success between educational levels and in groups with different amounts of training.

The reliability of the tests was described in five different ways: (1) objectivity; (2) length; (3) sampling of subject matter; (4) selection of the schools used in the testing; and (5) statistical measures, primarily a split-half count of alternate-items, corrected by the Spearman-Brown formula, yielding reliability coefficients for the different forms of the test and for the various grade and class years and for sex.

The study has shown, also, general differences in the ability to use the library as revealed by the sectional scores. This has emphasized the relatively greater ability to locate individual books than to use effectively the contents of reference books. It has shown the need for more detailed analysis of the problem of library skills. Such analysis should be based on the comparison of test questions, and on the correlation of the sectional scores with specific skill tests in subject fields.

